

Is There a Future for Cataloging?

Diane I. Hillmann

May 2006

In the last few months some heavy hitters have come forth with sets of predictions and recommendations focused on the current problems associated with catalogs, bibliographic access, and the future of traditional cataloging. In order of publication, these reports, by the University of California Libraries [1], Karen Calhoun for the Library of Congress [2], and Indiana University Libraries [3], have provided much fodder for discussion, concern, not to mention downright hysteria by bloggers, discussion list participants, and the librarian on the street. The general consensus is that we have been blessed with living in interesting times, and we'll need to change what we do in response to developments in the broader world--beyond that there are few clear agreements between the reports.

It's difficult to argue with the point that the current state of affairs cannot be maintained. Traditional cataloging worked pretty well when most cataloging effort was aimed at the print-based scholarly output shared in some measure by the majority of academic libraries, with resulting metadata designated for use within local catalogs and bibliographic utilities. 'Twas a thing of beauty in many respects, this nicely closed system developed over many decades and honed to relative sharpness and efficiency over time. (Yes, Virginia, there were gaps, problems, bits of silly tweakiness and obsessions with punctuation that drove a good many of us round the bend on a regular basis. Only when I wandered into the chaos of the metadata universe did I appreciate the shared cataloging beast properly, and pine again for those shelves full of citable documentation and the certainties of long experience.)

These three reports should be required reading for anyone who currently works in a library catalog department and is below the age of 60 or so. Among the general agreements between the reports is that libraries must move beyond MARC 21 and begin to work with other kinds of metadata, both as reusable data acquired from others and as data created in-house. Particularly as efforts to manage unique digital collections become more of a focus for libraries than managing the formerly more common print resources, non-MARC metadata will become the norm, even though many challenges of interoperability remain to be solved. This general agreement has, of course, major implications for the development of RDA, as mentioned in my previous comments in this space.

Although I suspect I'll come back again to dig in the rich loam of these reports, I'm going to focus in this column on the subject of ... subjects, or, more generally the organization of knowledge that is in my opinion the real value that libraries bring to the information universe. Clearly some bathwater will be tossed out as we figure out where we should be spending our precious limited resources, but we need to identify the baby before we start

the tossing. I'm not sure we've done that adequately yet, though each of the reports mentions subject access and comes to a slightly different conclusion.

Who's your baby?

In late April I made a brief visit to an old friend and colleague at the Library of Congress while in Washington on other business. I've known Jolande Goldberg (Senior Cataloging Policy Specialist, Cataloging Policy and Support Office) since I was a practicing law librarian, too many years ago to think about. Jolande has been responsible for the completion of the LC law classification schedules, and brings incredible energy and creativity to that effort. During my visit I was treated to a glimpse of her current work on the classification schedules for the law of indigenous peoples, and I was tremendously inspired by what I saw. What Jolande is creating is truly an exemplar of knowledge organization, using the bones of a fully researched and articulated classification schedule to organize and link to the increasing body of information available now on the web about legal issues and laws applying to native peoples of the world. Anyone applying classification numbers to a resource using this classification schedule will also be linking to this research. This is not your mama's "mark-it-and-park it," classification, but instead should be recognized as the newest node on the Library of Congress Classification System (LCC), a knowledge organization system truly awesome in its comprehensiveness and potential application in the digital world.

Those of you who think of classification as pages of printed schedules in bound volumes gathering dust on catalog department shelves are woefully behind the times. LCC has been fully encoded in the MARC 21 Classification Format for a number of years. At the moment those classification records are sadly underused, available primarily via a bespoke application developed at LC for the use of catalogers. And, of course, because most physical books carry only one classification number (and several subject headings), the potential of multiple classification numbers to expand access to increasingly multidisciplinary digital resources has been largely overlooked.

Classification is useful because it places information within a potentially browsable hierarchy of subject concepts, but unfortunately the terminology in the class schedules is limited and too often out of date. In addition, the wide use of tables to save paper (imported into the machine encoding to assist in the printing of the schedules) creates significant limits to the out-of-the-box usefulness of the encoded schedules as a tool for supporting browse access to collections. LCSH contains more updated terms, as well as thesaural "lead in" terms (sometimes the terms that are outdated but still used in the class hierarchies) to the current concepts and expresses relationships between terms, but there are no linkages between LCSH and LCC outside the LC ClassWeb product. Therefore, a subject strategy that makes use of these tremendous resources cannot really be completely realized by emphasizing one over the other.

So, let's go back to the question of identifying the baby (remember the baby?). When we talk about "cataloging" we tend to mash together description and access (including subject access), but in fact the determination of where the value lies in our common

legacy of bibliographic practice requires us to separate those discussions. There are especially good reasons to do that now, particularly as RDA has become the lightning rod for widespread concern about where we're headed in the descriptive realm.

Karen Calhoun in her report to LC talks about “dismantling” LCSH, but I've checked with her and confirmed that her intent is to support a more post-coordinated approach to subject access, de-emphasizing the classic pre-coordinated subject strings and using instead separate topic, genre/form, and geographic thesauri more amenable to simpler or automated application. Most librarians unembedded in the metadata world don't yet realize that there's been a tremendous explosion of interest in controlled vocabularies (not to mention the less formal “folksonomies”) in the world of the Web. Between LCSH, classification and the various authority files, the Library of Congress is sitting on the biggest, richest set of controlled vocabularies in the world, most of them created with the help of the broader library community, but in fact too often treated as a costly burden by the Library of Congress. This is increasingly seen as a problem, since LC's willingness to manage and make decisions about the future of these resources, in the community's interest rather than just LC's, is sometimes hampered by a lack of understanding of how to web-enable vocabularies, a disinclination to consult the library community on tough issues where they'd rather act unilaterally, coupled with an acknowledged lack of available resources and restrictive Congressional mandates.

It's my contention that it is these files and their embedded expertise that should be the real impetus to refocus the discussion around what knowledge organization means in our future world—and that it is this legacy that constitutes the real “baby” that must not be thrown out with the proverbial bathwater. Right now the vocabularies are available to libraries primarily through vendor applications, or via the Cataloging Distribution Services as files for sale. This severely limits their usefulness in the coming metadata world.

There have been small “baby” steps taken to explore the processes needed to make some of LC's smaller authority files available to web applications. The MARC Relator terms are now “webified” and available for use as roles within Dublin Core metadata (see: http://dublincore.org/documents/usageguide/appendix_roles.shtml for more information on how that works). But there's been little planning (at least that I've seen) at addressing the issue of making the other files available on a similar basis. What we hear too often is recommendations for saving money, not making the investments of a century of librarian effort and expertise available to us via the web—for us, in our new guise as metadata librarians, and for others only now discovering that metadata scales up well only with controlled terms as part of the picture.

Let's get talking.

[1] University of California Libraries. Bibliographic Services Task Force. Rethinking how we provide bibliographic services for the University of California. Final report, December 2005. Available at: <http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>

[2] The Changing Nature of the Catalog and Its Integration with Other Discovery Tools: Final Report, March 17, 2006, Prepared for the Library of Congress by Karen Calhoun. Available at: <http://www.loc.gov/catdir/calhoun-report-final.pdf>

[3] Indiana University Libraries Future of Cataloging Task Force, "A White Paper on the Future of Cataloging at Indiana University," Available at: http://www.iub.edu/~libtserv/pub/Future_of_Cataloging_White_Paper.doc